DOCUMENT RESUME

ED 353 560 CS 011 162

AUTHOR Hodges, Carol Ann
TITLE Literacy Assessment in Kindergarten: A Longitudinal Study of Teachers' Use of Alternative Forms of Assessment.
PUB DATE Dec 92
NOTE 19p.; Paper presented at the Annual Meeting of the National Reading Conference (42nd, San Antonio, TX, December 2-5, 1992).
PUB TYPE Speeches/Conference Papers (150) -- Reports - Research/Technical (143)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Elementary School Students; Elementary School Teachers; Evaluation Methods; *Kindergarten; Literacy; Longitudinal Studies; Primary Education; Reading Achievement; Reading Research; *Reading Tests; *Standardized Tests; *Student Evaluation; *Test Use; *Writing Tests
IDENTIFIERS *Alternative Assessment

ABSTRACT

A study compared teachers' assessments of students based on alternative evaluation techniques to student assessments based on standardized tests. Teachers in seven kindergarten classrooms evaluated their students (approximately 130 each year) in each of 3 successive years according to how well they had mastered a set of criteria which the teachers felt represented the successful reader and writer at the end of kindergarten. Standardized tests were also administered to the kindergarten students. Results indicated that: (1) a significant relationship existed between teachers' assessments of students and the students' performance on the standardized test; (2) 56% of the variance between the total teacher groupings and the total test scores appeared to be due to some common factors; (3) when results differed, teachers ranked students in the next lower category 76% of the time the first year, 88% of the time the second year, and 98% of the time the third year. Follow-up interviews indicated that teachers felt more confident in their ability to make decisions about students' abilities; parents and teachers felt that teachers' evaluations provided more useful information than the standardized tests did; the principal did not agree to ban standardized tests as the teachers had requested. Findings suggest that teacher judgments, based on knowledge of their students' development and knowledge of the processes involved in reading and writing, may be a more valid means of obtaining information for instructional decisions. (One table of data and the kindergarten reading strategies checklist are included; 16 references are attached.) (RS)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED353560

Carol Ann Hodges
Buffalo State College
25 Hampton Hill Lane
Williamsville,NY 14221
(716) 633-7290

Literacy Assessment in Kindergarten:

A Longitudinal Study of Teachers' Use of Alternative Forms of Assessment

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Carol A. Hodges
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

CS011163

Running head: ALTERNATIVE ASSESSMENT IN KINDERGARTEN

BEST COPY AVAILABLE

Literacy Assessment in Kindergarten: A Longitudinal Study of Teachers' Use of Alternative Forms of Assessment

The use of standardized tests has increased dramatically over the past few decades and the trend toward more testing seems likely to continue. However, as the emphasis on standardized tests has escalated, so have objections to them. A number of reading researchers (Edelsky & Harmon, 1988; Garcia & Pearson, 1991; Hodges, 1991, 1992; Johnston, 1992; Morrow & Smith, 1990; Squires, 1987; Teale, 1988, Valencia & Pearson, 1986) have pointed out that early reading assessment has not kept pace with advances in reading research, theory, and practice. At the same time early childhood experts (Bredekamp, 1986; Fairtest & NYPIRG, 1990; Harmon, 1990; International Reading Association, 1986; Kamii, 1990; Moyer, Egertson, & Isenberg, 1987; National Association for the Education of Young Children, 1988) argue that children are being tested too early. They claim that young children are not good test takers; that the unfamiliar format leads to stress; that test results are influenced by the children's ability to sit still and be quiet; and that extensive testing narrows and misdirects the curriculum and drains instructional time without a clear demonstration that the investment is beneficial. In addition, groups as diverse as the American Association of Colleges for Teacher Education (AACTE), the American Federation of Teachers (AFT), the National Association of Elementary School Principals (NAESP), and the national PTA have spoken out to urge states to abandon the use of multiple-choice tests and to replace them with alternative assessment techniques which seek to measure directly the student's ability to perform in the subject area.

of efficiency and objectivity. Because it is easy to get a false sense of security when skilled reading is equated with scores on reading tests, many school personnel and parents continue to believe that data from standardized tests are more trustworthy than data collected by other means.

PROBLEM

After completing the survey of early reading tests, I began investigating the primary level (K-2 grade) literacy program and assessment tools of a school district in a small suburban/rural community in the eastern part of the United States. This district tested all of its students beginning in kindergarten each May with a widely used standardized test battery. After interviewing the kindergarten teachers I discovered that they administered standardized reading achievement tests to students very reluctantly. They resented the time that the administration of the test took from instruction, the pressures that it put on the curriculum, and the frustration that it exerted on their students. In addition, because these teachers were making the transition from a basal readiness program to a more developmentally based process oriented literacy program, they felt the need to have a variety of assessment tools for the everyday instructional decision-making that is a crucial part of that approach. But they were not sure how to use informal assessment and, even they wondered whether the informal tools could provide valid and reliable data.

The questions most often asked by the teachers, the administrators, and the parents were, "How can teachers use alternative evaluation techniques?" "How do teachers' assessment of students based on alternative evaluation techniques compare to the way in which the standardized test assesses them?

How valid are teacher judgments? The results of this study provide some information to answer those questions.

METHOD

Teacher Ratings: Year One

During the first year of the study, before they administered the standardized achievement test, I asked the teachers in the seven kindergarten classrooms to evaluate their 136 students according to how well they had mastered a set of criteria which the teachers felt represented the successful reader and writer at the end of kindergarten. Among the criteria reported by the teachers were the following: (a) the students' attitude toward books and reading/writing, (b) their recognition of the letters of the alphabet, (c) their knowledge of grapheme/phoneme correspondences, (d) their ability to listen to and comprehend stories, (e) their ability to read independently, and (f) general maturity, a concept which the teachers further defined as following directions and keeping to a task.

The teachers assigned their students a score of (3) if they were above average readers/writers, a (2) if they were average readers/writers, and a (1) if they were below average readers/writers based on the aforementioned criteria that they observed in classroom behavior or in finished products. The teachers were also beginning to consider such criteria as knowledge of selected concepts of print, use of invented spelling in writing, and the ability to re-tell their written stories as important variables; but they did not feel secure in their ability to judge their students in these areas.

The standardized test which the teachers later administered to their kindergartners purported to assess skills in auditory discrimination,

grapheme/phoneme correspondence, decoding, and listening comprehension. The test scores used in the correlations were the Total Reading stanine scores (9-1) that the students earned. The degree of the relationship between the teacher ranked groups and the test scores was computed by using the Pearson Product Moment correlation coefficient.

Teacher Ratings: Years Two and Three

Over the next two years these kindergarten teachers met with me and attended a local kindergarten whole language support group (KTT-Kindergarten Teachers Together). They read widely in the field to broaden the theoretical framework underlying their instruction, to explicate the goals of that instruction, to clarify the purposes of their authentic assessment, and to devise the tools that they believed would be most appropriate for their purposes. These tools in their final formats were observation checklists, anecdotal records, and portfolios of childrens' work. Figure One is an example of one of the observational checklists which the teachers devised. As they used these informal measures of reading and writing abilities, they became more secure with their ability to make judgments about such variables as their students' knowledge of selected concepts of print, their use of invented spelling in writing, and their ability to retell their written stories.

Insert Figure One about here

Before the spring reading achievement test was administered each of those next two years, the teachers again assessed their students as being above

average readers/ writers (3), average readers/writers (2), or below average readers/writers (1) (Year Two, n=131; Year Three, n=125). The criteria reported during the first year's judgments were used as well as other criteria on which they had collected information as they used their new informal assessment tools. As before, the degree of the relationship between the teacher ranked groups and the standardized test scores was computed by using the Pearson Product Moment correlation coefficient.

In addition, during those years the kindergarten teachers, a group of parents, the first grade teachers, and the school principal were interviewed about their use of the results of both the standardized test data and the authentic assessment data for making instructional and policy decisions.

RESULTS AND CONCLUSIONS

Correlations

The first year a comparison of the teacher assessments with the Total Reading stanines reported on the standardized test showed that there was a significant relationship between the assessments of the students by the teachers and the Total Reading stanines obtained by the students on the standardized test. Table One illustrates that the correlations for the classes ranged from .59-.87 ($p<.01$). A correlation of .75 ($p<.01$) was found over all classes.

Insert Table One about here

Results for years two and three can also be found in Table One which illustrates that during the second year the correlations for the classes

ranged from .49-.89 ($p<.01$) with a correlation of .77 ($p<.01$) over all classes. Correlations for the third year ranged from .51-.87 ($p<.01$) with a correlation of .70 ($p<.01$) over all classes.

The coefficient of determination (r^2) for the first year's entire set of classroom groupings was equal to .56. Thus, over fifty percent of the variance between the total teacher groupings and the total test scores appeared to be due to some common factors. This implies that the teachers and the test were tapping different factors for the other 44 percent. Similar results were found over the second (r^2 =.59 and third years (r^2 =.49).

A closer look at the data from individual classrooms reveals that when there were differences between the teachers' rankings and the total test score obtained on the achievement test, teachers were more inclined to rank the students in a lower category than the test did. Results from the first year illustrate that when results differed, teachers ranked students in the next lower category 76% of the time, during the second year 88% of the time, and during the third year 98% of the time.

At least three factors can help explain these differences. First, as the teachers devised and became more comfortable with the use of their own informal assessment measures, they began to consider more variables in their judgments than the test did. While the students who were judged higher by the test may have been successful in auditory discrimination or decoding skills, the teachers did not believe that those same students were as successful in their knowledge of selected concepts of print, use of invented spelling in writing, ability to retell written stories, ability to read independently, and their attitude toward books and reading/writing.

Second, while the teachers in the school district where the research project took place were accustomed to judging their students in terms of the local children they had taught in previous years, those same students traditionally ranked above the average on the test in terms of national norms. Thus, some students who were considered average by national standards might be ranked below average by their teachers who made comparisons based on their past experiences with a generally above average student population.

Third, because teachers were asked to evaluate their students as above average, average, or below average, they were, in a sense, predisposed to categorize some pupils in each class as below average. Therefore, in some classes no children received a below average stanine test score (1-3), but did receive a below average assessment by the teacher. Any replication of this study should word the directions to teachers carefully so that they do not feel pressured to place students in a below average category.

Another word of caution for teachers must be added at this point. In some cases students who were placed in the belcw average category but who had received higher scores on the test were judged below average by teachers primarily because of their lack of maturity. While it is impossible to define exactly what the term means, it was evident from some teacher comments that lack of maturity was at times synonymous with discipline problems. Might discipline problems be a signal that a child is bored and possibly misdiagnosed? Teachers who use informal measures to judge student ability must be sure that students with discipline problems are not judged to be below average solely for that reason.

Interviews

During follow-up interviews the kindergarten teachers reported that, based on the correlational evidence and the actual progress of their students over the three years, they became much more confident in their ability to make decisions about their students' reading and writing knowledge and ability. In addition, they stated that they used the informal measures not only for summative evaluations but also for the formative evaluations that guided their everyday instructional decisions. Parents and first grade teachers believed that the kindergarten teachers' judgments, based upon multiple authentic measures, provided more useful information at conferences and in end of the year reports than the test data did. Some parents, however, still felt that standardized test scores provided important and necessary information, even though they were not sure what that information meant. The principal, while accepting the authentic measures as valid and reliable indicators of the students' reading and writing ability and admitting that the test scores were rarely used for instructional purposes, has not been convinced to support the kindergarten teachers' request to ban standardized testing in their kindergarten classrooms.

IMPLICATIONS

What evidence would prove that teacher judgments can be valid measures of reading/writing achievement? If we were to develop a new traditional test of reading/writing achievement, we would have to find a valid criterion measure of reading/writing to establish the new test's concurrent validity. Because we know that there are no perfect measures of reading/writing achievement, we would probably use other reading achievement tests that are presumed to be valid. Then if our new test elicited test scores correlating significantly

with the other tests, we would conclude that our new test was a valid measure of reading achievement. Why shouldn't we in this case use the correlations found between the teacher assessments and the test scores to establish concurrent validity?

The question may really be, "Do we want to"? First, can we presume that the standardized test used by the school district in this study is a valid one? The technical manual of the test used states that the test is *expected* to correlate significantly with other achievement measures but offers no specific data to support the claim. And how do we know that the other tests are valid measures? As has already been stated, most reading assessment has not kept pace with advances in reading research, theory, and practice. And, even if this particular test correlated highly with other similar tests, would it necessarily be a valid test of reading/writing as they are conceived of in this school district?

This is an important question. The current debate over national standards has raised a number of perplexing issues concerning just how we are to come to agreement on those standards, and how we are to assess them. It would be a travesty if any school simply relinquished its responsibility for the education of its children to a standardized test that may be based on conceptions of reading/writing at odds with either national or local conceptions or both. At a bare minimum the district would need to articulate the conceptions of reading/writing in which it believes and then determine whether the existing tests conform to those conceptions or not.

Furthermore, since "validity" applies not to tests but to the <u>inferences</u> we make from those tests, an important question to ask would be,

"What kinds of information are used in the school district when decisions are being made?" First grade teachers and the school principal who were interviewed about their use of previous end-of-the-year test results and assessments made by their students' previous teachers unanimously chose to use the previous teachers' assessments over the test results. By their actions they show that more of what they believe is truly important in reading/writing is captured by the kindergarten teachers' assessments than by the standardized test. This fact can be explained in one of two ways. Either the principal and teachers are making persistent mistakes in not trusting an "objective" test, or the teachers' assessments are, indeed, a more adequate means of measuring the reading/writing abilities of the children. Since these assessments appear both to capture many of the same things as the standardized test and to go further in picking up on the features important to real instructional decisions, the automatic suspicion of teacher judgments appears itself to be highly suspect.

Of course, the results of this study are limited because the population consisted of only one school district. However, having found such consistency of medium to high correlations, I believe that the teacher and the test measures are likely measuring a number of similar factors. The coefficient of determination over the three years ranged from .49-.59, leading me to believe that the teachers and the test were tapping nearly fifty percent of the same factors. The relatively high correlations of teacher judgment with standardized tests should ease fears that teacher judgments would be totally at odds with the standardized test results.

Thus, knowing what we do about the negative factors associated with standardized tests and testing in the primary grades and the fact that little use seems to be made of the test results, the data suggest that teacher judgments, based on knowledge of their students' development and knowledge of the processes involved in reading and writing, may be even more valid means of obtaining information for instructional decisions. I urge others to replicate this study. If pupil assessments by teachers in other school districts also correlate moderately highly with test scores and are used more regularly for instructional decisions, then the notion of "subjectivity" in the alternative forms may not be the negative factor that some now consider it.

References

Bredekamp, S. (Ed.). (1986). Developmentally appropriate practice. Washington, D.C.:National Association for the Education of Young Children.

Edelsky, C. & Harman,S. (1988). One more critique of reading tests-with two differences, English Education 20, 157-71.

FairTest & NYPIRC. (1990). Standardized tests and our children: A guide to testing reform in New York. New York: The National Center for Fair and Open Testing.

Garcia, G. & Pearson, P. D. (1991). The role of assessment in a diverse society. In E. Hiebert (Ed.), Literacy for a diverse society:Perspectives, practices, and policies. (pp. 253-278), New York: Teachers College Press.

Harmon, S. (1990). Negative effects of Achievement testing in literacy development. In C. Kamii (Ed.), Achievement testing in the early grades. (pp. 111-118), Washington, D.C.: National Association for the Education of Young Children.

Hodges, C. (1991). Instruction and assessment of emergent literacy. In L. Weis, P. Altbach, G. Kelly, & H. Petrie (Eds.), Critical Perspectives on early childhood education. (pp 153-168), New York: State University of New York Press.

Hodges, C. (1992, April) Teacher judgments and standardized assessments. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.

International Reading Association, Early Childhood and Literacy Development Committee.(1986). Literacy development and pre-first grade.Childhood Education. 63, 110-11.

BEST COPY AVAILABLE

Johnston, P. (1992). Constructive Evaluation of literate activity. New York: Longman Publishing Group.

Kamii, C. (1990) Achievement testing in the early grades: The games grown-ups play. Washington, D.C.: National Association for the Education of Young Children.

Morrow, L. & Smith J. (1990) Assessment for instruction in early literacy. Englewood Cliffs, NJ: Prentice-Hall.

Moyer, J., Egertson, H., & Isenberg, J. (1987). Association for childhood education international position paper: The child-centered kindergarten, Childhood Education. 63, 235-42.

National Association for the Education of Young Children. (1988). NAEYC position statement on standardized testing of young children 3 through 8 years of age. Young Children 43, 42-47.

Squires, J. (1987, April). Introduction: A special issue on the state of assessment in reading. The Reading Teacher, 40 , 724-725.

Teale, W. (1989). Developing appropriate assessment of reading and writing in the early childhood classroom. The Elementary School Journal 89, 173-183.

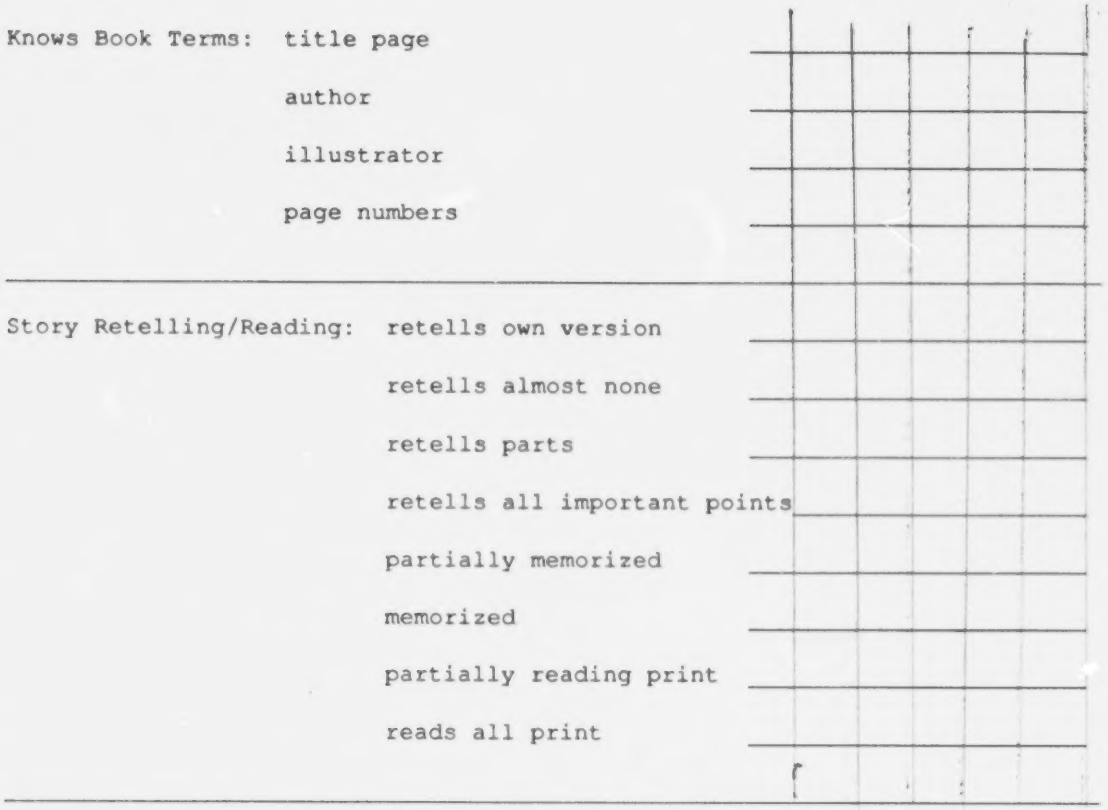
Valencia, S. & Pearson P. D. (1986). New models for reading assessment. Reading Education Report 71. Urbana, IL: University of Illinois, Center for the Study of Reading

Figure Caption

Figure 1. Kindergarten reading strategies checklist

NAME:_____

Coding: + = knows, - = learning	DATES:				
Identifies front of book					
Knows where to start reading					
Aware of page turning direction					
Aware of top-bottom reading					
Aware of left-right					
Aware of return sweep					
Knows punctuation: period					
question-mark					
exclamation-point					
other					
Can identify a letter					
Can identify a word					
Knows print contains message					
Finger pointing: no attempt					
word by word					
slides across					
Knows Book Terms: cover					
title					



Knows g/p correspondence (circle)

b c d f g h j k l m n p q r s t v w x y z a e i o u

Sight words and notes:

Table 1

Pearson Product Moment Correlations* of Teacher Assesments Standardized Test Scores

Kindergarten Classrooms	Year One Correlations	Year Two Correlations	Year Three Correlations
1	.61	.89	.51
2	.86	.80	.83
3	.72	.75	.86
4	.87	.84	.69
5	.85	.78	.87
6	.72	.49	.43
7	.59	.83	.69
All kdgs.	.75	.77	.70

*p<.01